NL 010043
US
NL 010450
I

Europäisches Patentamt	European Patent Office	Office européen des brevets

Bescheinigung | Certificate | Attestation

Die angehefteten Unterlagen stimmen mit der ursprünglich eingereichten Fassung der auf dem nächsten Blatt bezeichneten europäischen Patentanmeldung überein.

The attached documents are exact copies of the European patent application described on the following page, as originally filed.

Les documents fixés à cette attestation sont conformes à la version initialement déposée de la demande de brevet européen spécifiée à la page suivante.

Patentanmeldung Nr. | Patent application No. | Demande de brevet n°

01200144.2

Der Präsident des Europäischen Patentamts;
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets
p.o.

I.L.C. HATTEN-HECKMAN

DEN HAAG, DEN
THE HAGUE, 25/10/01
LA HAYE, LE

Europäisches Patentamt
European Patent Office
Office européen des brevets

Blatt 2 der Bescheinigung
Sheet 2 of the certificate
Page 2 de l'attestation

Anmeldung Nr.:
Application no.: 01200144.2
Demande n°:

Anmeldetag:
Date of filing: 16/01/01
Date de dépôt:

Anmelder:
Applicant(s):
Demandeur(s):
Koninklijke Philips Electronics N.V.
5621 BA Eindhoven
NETHERLANDS

Bezeichnung der Erfindung:
Title of the invention:
Titre de l'invention:
Sinusoidal linking mechanism

In Anspruch genommene Priorität(en) / Priority(ies) claimed / Priorité(s) revendiquée(s)

Staat: State: Pays:	Tag: Date: Date:	Aktenzeichen: File no. Numéro de dépôt:

Internationale Patentklassifikation:
International Patent classification:
Classification internationale des brevets:
/

Am Anmeldetag benannte Vertragstaaten:
Contracting states designated at date of filing: AT/BE/CH/CY/DE/DK/ES/FI/FR/GB/GR/IE/IT/LI/LU/MC/NL/PT/SE/TR
Etats contractants désignés lors du depôt:

Bemerkungen:
Remarks:
Remarques:

Sinusoidal linking mechanism

A sinusoidal linking mechanism based on similarities of the constituent complex signals at the seam

A.C. den Brinker, A.W.J. Oomen, F.M.J. de Bont and E. Schuijers

In sinusoidal coding, linking of sinusoidal parameters from subsequent frames is used in order to reduce redundancy. Typically this means differential coding of sinusoidal parameters and/or removal of the phase information along tracks.

A proposal for a linking criterion is made which is based on an error measure of the instantaneous signal parameters at the seam. This proposal can be applied to currently investigated extensions of sinusoidal track models.

Things known so far

In current sinusoidal coders, sinusoidal tracks are formed by estimation of sinusoidal parameters per frame (segment) and a linking procedure which establishes a similarity between estimated signal components in two subsequent frames.

The linking criterion has changed over time. The first proposal concerned the link based on a frequency distance [1, 2]. Later, relative frequency distance were proposed as the basis for a criterion [3] and combination of relative distances in frequency and amplitude as well [4]. We note that none of these methods use the phase information.

A more elaborate way of establishing tracks was proposed in WO 00/79519-A1 were partial signals are reconstructed based on all possible parameter links and these are compared with the original signal. The drawback of this algorithm is it computational burden but, in contrast to the previous mechanisms, it does take the phase information into account.

Extensions of the sinusoidal model have been proposed. Next to the estimation of frequency, amplitude and phase (or derivatives thereof) other free parameters are involved, see [5, 6, 7] and the co-pending European patent application entitled "sinusoidal matching pursuit", filed by the Applicant on the same day as this application.

The problem for which this invention brings the solution

All the above mentioned methods have problems. As already mentioned, the first set of methods [1, 4, 3] do not use phase information and may therefore provide incorrect tracks. The second method (*WO 00/79519*) is computationally expensive. Lastly, linking mechanisms for extended sinusoidal models have not been proposed yet.

A linking mechanism is proposed which uses phase information, is computationally not as expensive as *WO 00/79519* and which can also be applied for extensions of the sinusoidal model. The basic idea is to consider the constituent complex signals of two subsequent frames and base the similarity measure on the instantaneous signal values and instantaneous stride of these complex signals at the seam.

Embodiment

In sinusoidal modelling, the models are typically of the form (or can be rewritten as such)

$$s(t) = \sum_{k=1}^{K} \Re\{u_k(t)\} \tag{1}$$

where u_k are the underlying sinusoidal or sinusoidal-like signals.

If we consider two subsequent signal segments, s_1 and s_2, then there is typically overlap in their support. In order that profitable (in a coding sense) links are established, it seems reasonable to speak of a link between a component m from s_1 and n from s_2 only if $u_m(t)$ and $u_n(t)$ are similar within the overlap area.

Method 1

First, consider the complete overlap. The aim is to identify signals in this overlap which are similar. This can be done by a correlation method. We define the correlation coefficient $\rho_{m,n}$ by

$$\rho_{m,n} = \frac{\sum_t w(t) x_m(t) y_n^*(t)}{\sqrt{E_{x_m} E_{y_n}}} \tag{2}$$

where x_m ($m = [1, M]$) and y_n ($n = [1, N]$) are the sets of function we want to consider, w is a window function, E_v is the energy in signal v

$$E_v = \sum_t w(t) v(t) v^*(t) \tag{3}$$

ρ is a complex value which. for a link, should be close to 1. Therefore, we can build a (partial) similarity measure as before. Thus,

$$S_1(m,n) = \begin{cases} 1 - |\rho_{m,n} - 1| / D_1, & \text{if } |\rho_{m,n} - 1| < D_1, \\ 0, & \text{elsewhere,} \end{cases} \tag{4}$$

with $0 < D_1 < 1$.

Additionally, the equivalence in amplitude (or, more particular, in energy) can be taken into account by considering

$$R_{m,n} = \min\left\{\frac{E_{x_m}}{E_{y_n}}, \frac{E_{y_n}}{E_{x_m}}\right\}. \tag{5}$$

Again, for a link, R should be a value close to 1 (but now real-valued) and we propose as similarity measure

$$S_2(m,n) = \begin{cases} 1 - (1 - R_{m,n})/D_2, & \text{if } (1 - R_{m,n}) < D_2, \\ 0, & \text{elsewhere,} \end{cases} \tag{6}$$

with $0 < D_2 < 1$.

As an overall similarity measure this leads to

$$S(m,n) = S_1(m,n)S_2(m,n), \tag{7}$$

where $S(m,n) = 0$ means that there is no link and the larger $S(m,n)$ is, the more likely it is that this can be exploited profitably as a link in a sinusoidal coding scheme.

If signal s_1 is approximated by M components and s_2 by N, we can construct an $M \times N$ matrix and from the entries establish if there exist links and the most profitable ones.

Method 2

To simplify the procedure outlined above, we look within the overlap region at the middle specifically. At this point (let's say t_0) we should have

$$u_m(t_0) \approx u_n(t_0). \tag{8}$$

We would like that in the the neighbourhood of t_0 the signals match as well. This is realised if the progression (the stride) in the signals is (nearly) the same e.g., evaluated by

$$\frac{u_m(t_0+1)}{u_m(t_0)} \approx \frac{u_n(t_0+1)}{u_n(t_0)}. \tag{9}$$

In order to select links we now propose the (partial) similarity measure

$$S_3(m,n) = \begin{cases} 1 - \left|\frac{u_m(t_0)}{u_n(t_0)} - 1\right|/D_3, & \text{if } \left|\frac{u_m(t_0)}{u_n(t_0)} - 1\right| < D_3, \\ 0, & \text{elsewhere,} \end{cases} \tag{10}$$

with $0 < D_3 < 1$. We note that the amplitude similarity is involved in a relative way. This agrees with psycho-acoustic relevance and distance criteria.

The second partial similarity measure is defined as

$$S_4(m,n) = \begin{cases} 1 - \left|\frac{u_m(t_0+1)}{u_m(t_0)}\frac{u_n(t_0)}{u_n(t_0+1)} - 1\right|/D_4, & \text{if } \left|\frac{u_m(t_0+1)}{u_m(t_0)}\frac{u_n(t_0)}{u_n(t_0+1)} - 1\right| < D_4, \\ 0, & \text{elsewhere,} \end{cases} \tag{11}$$

with $0 < D_4 < 1$.

The total similarity measure is defined as

$$S(m,n) = S_3(m,n)S_4(m,n). \quad (12)$$

It is obvious that instead of looking at (relative) differences between complex values u, we can also look at the real and imaginairy part or amplitude and phase of u and construct the similarity criterion. This has the advantage that instead of the two parameters that control the above given similarity measure, we get one or more parameter per considered variable. Therefore, expressed in real parameters instead of complex ones, we typically end up with twice as many parameters.

Application areas

Audio and speech coding.

References

[1] R. McAulay and T. Quartieri. Speech analysis/synthesis based on sinusoidal representation. *IEEE Trans. Acoust., Speech, Signal Process.*, 43:744–754, 1986.

[2] R.J. McAulay and T.F. Quartieri Jr. Processing of acoustic waveforms. US Patent 4,885,790, Dec. 5, 1989.

[3] S.N. Levine. *Audio representation for data compression and compressed domain processing.* PhD thesis, Stanford Univ. (CA), 1999. Pages 1–136.

[4] B. Edler, H. Purnhagen, and C. Ferekidis. ASAC – Analysis/synthesis codec for very low bit rates. Preprint 4179 (F-6) 100th AES Convention, Copenhagen, 11-14 May 1996.

[5] E.B. George and M.J.T. Smith. A new speech coding model based on a least-squares sinusoidal representation. In *Proc. 1987 Int. Conf. Acoust. Speech Signal Process. (ICASSP87)*, pages 1641–1644, Dallas TX, 6-9 April 1987. IEEE, Picataway, NJ.

[6] M.M. Goodwin. *Adaptive signal models: theory, algorithms, and audio applications.* PhD thesis, Univ. of California, Berkeley, 1997. Pages 1–259.

[7] J. Nieuwenhuijse, R. Heusdens, and E.F. Deprettere. Robust exponential modeling of audio signals. In *Proc. SPS 98*, pages 143–146, Leuven, Belgium, 26–27 March 1998.

In the following, an embodiment of the invention is described.

SinuSoidal Coding (SSC) is a parametric audio coding technique, developed at Philips Research, aimed at a bit-rate of approximately 40 kbit/s for high quality stereo audio. The SSC coder divides audio into three objects: transients, sinusoids and noise. For each object relevant parameters are extracted and efficiently encoded into a bit-stream.

One of the most important steps for the compression is the use of the tracking algorithm. The tracks formed by this algorithm can be encoded very efficiently using differential encoding. Furthermore, phaseless reconstruction can be applied to further improve coding gain. As this is not a lossless process, the course of the tracks actually determines the quality. It is shown how, using the information of 2nd order polynomials, the linking of individual sinusoids is improved.

Furthermore, in the algorithm that removes the non-tracks a psycho-acoustic model is applied to further increase the audio quality.

AUDIO CODING

Audio coding is the process of encoding/decoding digitised audio signals. This is done in such a manner that after encoding the data rate is kept as low as possible while maintaining as much as possible of the original quality after decoding. Mainly thanks to the Internet, audio coding is also publicly known, in particular the MPEG-1 Layer III standard, better known as MP3. This compression scheme delivers high quality audio at compression gains of over a factor of 10. Since the standardisation of MPEG-1 many new and innovative ideas have been brought forward. In practice, this led to state-of-the-art MPEG4-AAC coders that provide compression-factors of around 15 while still maintaining the same high quality level as with MPEG-1 Layer III. The current opinion in the audio coding community is that no further improvement in compression gains is expected for waveform type of coders. There is a general belief that, in order to achieve even higher compression gains, audio should be coded parametrically.

PARAMETRIC CODING

When the input signals are restricted (e.g. to speech only) specific features of the input signal can be exploited to further improve the coding gain. One way of implementing these features in an audio coder is to use a parametric coding scheme. The main difference with a waveform coder is the explicit usage of a source model. For e.g. a speech model this is based on the human vocal tract. For a parametric model of a piano the hammers, the snares and the cabinet could be considered. Figure 1 depicts a schematic description of a parametric audio coder.

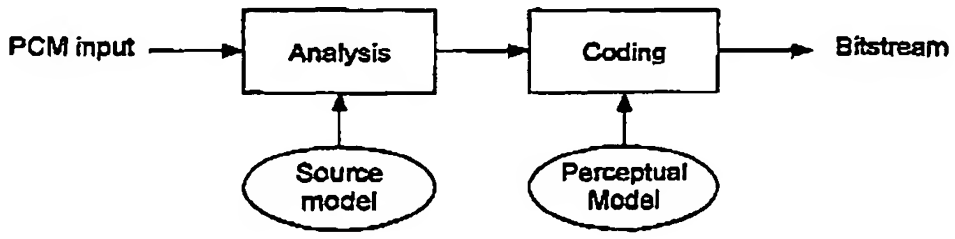

Figure 1: Parametric audio encoder

The main advantage of parametric audio coding is the exploitation of both the sender-end (source of sound) as well as the receiver-end (human auditory system), where the waveform coder exploits the receiver-end only. If the input signal however doesn't appropriately fit the source model this might lead to unpredictable results. This lack of robustness then again is the main disadvantage of parametric audio coding.

Another advantage of parametric audio coding is the perceptual model. This model can be adapted fully to the source model. If e.g. one object of the source model is defined as a set of harmonics, perceptual experiments using harmonics could be performed in order to come to a psycho-acoustic model for harmonics only.

Parametric models are also often called object-oriented models. Objects can be a bit abstract like for speech: 'tonal elements' and 'non-tonal elements' but also concrete like 'snare generated harmonics' of a guitar model. A description in terms of different objects automatically implies that within the encoder, decisions will have to be made about what part of the signal must be assigned to what object. The more such decisions have to be made the worse the robustness will probably be.

All in all parametric coding has more potential than waveform coding as far as bit-rate versus perceived audio quality is concerned. But, especially when designing a parametric model for a broader type of input signals, one has to consider robustness versus coding efficiency.

SINUSOIDAL CODING

The SSC coder is a parametric audio coder that aims at coding of audio in a broad sense, i.e. unrestricted to a classified source like e.g. speech. Therefore a source model must be developed that is both simple and effective. We have now classified audio into three objects, namely sinusoids, transients and noise. Such a description seems to be both simple as well as complete.

Another reason for choosing these objects as the base for a parametric audio model is the apparent relation to psycho-acoustics. When e.g. calculating the masking curve in most perceptual models of waveform coders a distinction is made between tonal elements (sinusoids) and non-tonal elements (noise). The masking effect of sinusoidal tones is different from the masking effect of noise.

The SSC codec is based on the three objects described above, namely: transients, sinusoids and noise. A block-diagram of the SSC-encoder is shown in Figure 2.

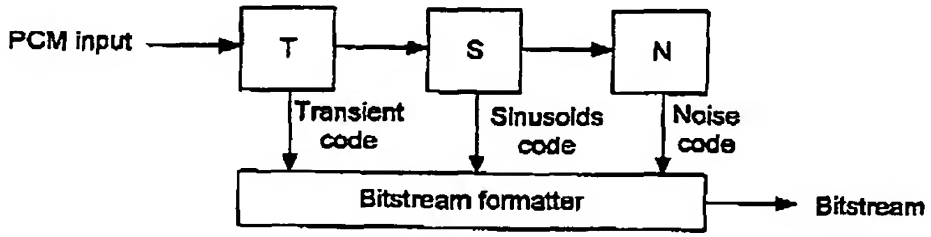

Figure 2: Block diagram of the SSC encoder

One of the most crucial steps in parametric audio coding is the subdivision of different parts of the signal to different objects. For both the analysis of sinusoidal components as well

as for analysing noise, quasi-stationarity is a prerequisite. So if transient phenomena could be removed first the residual signal will be more stationary and thus easier to analyse. This is the main reason why the transients module (T) has been placed in front of the other two. The sinusoidal module (S) is placed before the noise module (N) because it is much harder to analyse and remove the noise from the residual signal of the transient module than it is the other way around.

SINUSOIDS

The SSC project is based on the assumption that any digital audio signal can be described adequately by (Equation 13):

$$x[n] = \sum \text{transients} + \sum \text{sinusoids} + \sum \text{noise}. \qquad 13$$

It is however not clearly defined what a sinusoid is. Within the SSC project elements that are classified as being a sinusoid can be described by (Equation 14):

$$s[n] = \sum_p A_p(n)\cos\left(\omega_p(n)n + \varphi_p\right), \qquad 14$$

where $A_p(n)$ is the slowly varying amplitude, $\omega_p(n)$ is the slowly varying frequency and φ_p the phase of the p^{th} sinusoid. This representation was first used by McAulay and Quatieri for describing speech signals [McAulay and Quatieri 1986].

It is neither efficient nor feasible because of complexity to extract $A_p(n)$, $\omega_p(n)$ and φ_p on a sample-by-sample basis. A more feasible method would be to extract these parameters on a frame-to-frame basis. So for a single frame the sinusoids could be described by:

$$s_k[n] = \sum_p A_{p,k}\cos\left(\omega_{p,k}n + \varphi_{p,k}\right), \qquad 15$$

where k indexes the frame and p the p^{th} sinusoid.

In order to come to such a description another segmentation takes place on the PCM input signal. At this segmentation stage the space between two transient-positions is divided into overlapping frames of 720 samples. This number has been determined experimentally as a balance between stationarity and efficient coding of parameters. The segmentation is illustrated in Figure 3. The upper line describes the transient positions extracted by the transients module. The small blocks at the lower end show how the sinusoids are segmented between these transient positions. It is noted that the last frame of a segment is always placed in such a way that it ends just before another segment starts. This is also because of stationarity as described before by the first transient property.

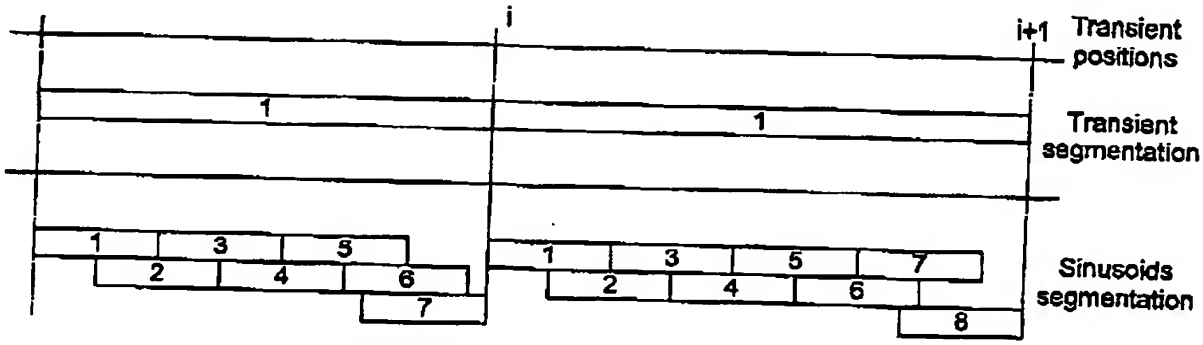

Figure 3: Segmentation of transients and sinusoidal module

The sinusoidal module can now be described as in Figure 4. It first of all consists of a sinusoidal analysis block (SA). In this block the sinusoidal parameters are estimated on a frame-to-frame basis. The sinusoidal synthesis block (SS) synthesises the sinusoidal signal from these parameters. Finally this signal is subtracted from the residual signal of the transient module to create a presumably noisy signal for the noise module.

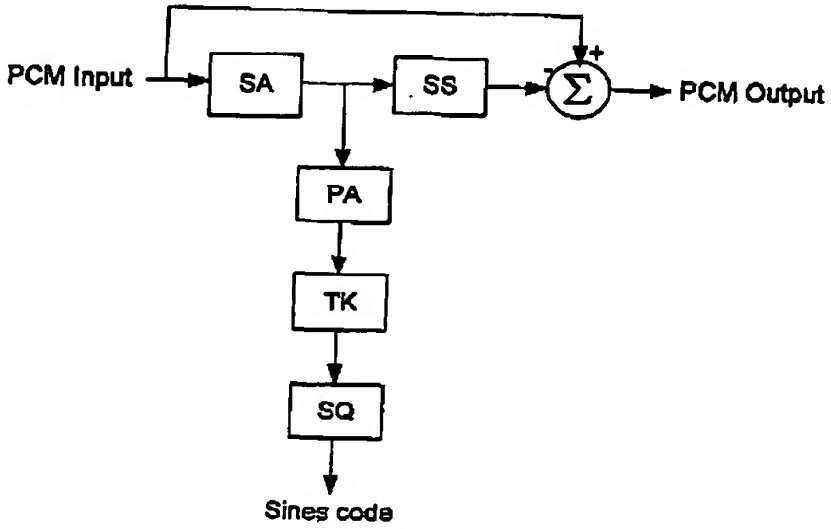

Figure 4: Block-diagram of sinusoidal module

Of course not all sinusoidal components that have been extracted are perceptually relevant. Inclusion of such components in the bit-stream is superfluous. Therefore a psycho-acoustic model (PA) is used to remove sinusoidal components that fall well below the masking threshold. The reason that the psycho-acoustic model is not used very tightly lies within the tracking block (TK).

To come to an efficient representation of all the individual sinusoidal components found over all the analysed frames a tracking algorithm (TK) is used. The main idea behind this algorithm is that sinusoids in general last longer than only a single frame and can thus form tracks. Differential encoding of e.g. amplitude and frequency can then prove to be efficient. This differential coding is the reason that the restraints of the psycho-acoustic model are set loosely. It is more efficient to code a long track differentially, even though it can't be perceived during the whole length of the track, than only encode the short relevant parts.

Even more coding gain can be achieved by applying phaseless reconstruction. This means that instead of updating the phase of a track at each frame only the phase of the birth of a track is encoded. For all following frames of the track the phase is calculated based on the presumption that the instantaneous frequency is a smooth and slowly-varying function in time.

Finally the sinusoidal quantisation (SQ) block codes the processed parameters to further increase coding gain. It does so by quantisation and coding of the parameters extracted for frequency, amplitude and phase.

Sinusoidal Analysis

The sinusoidal analysis block consists of an iterative algorithm for extracting the sinusoidal parameters . The block diagram is depicted in Figure 5.

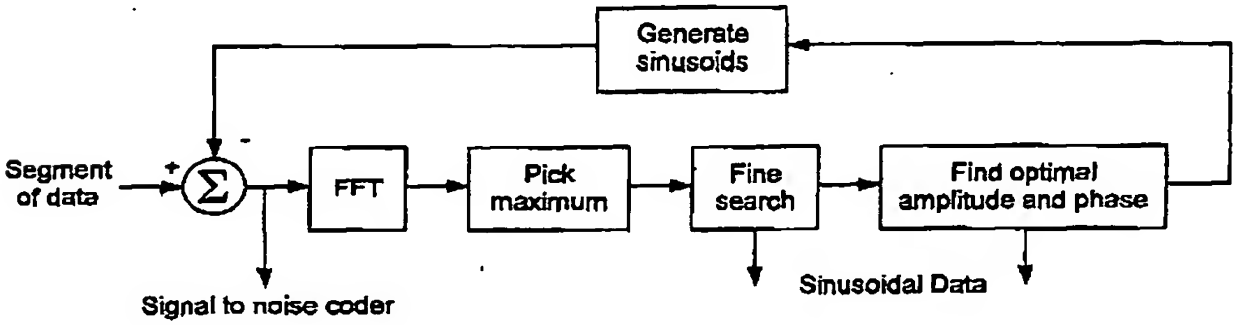

Figure 5: Extraction of sinusoidal parameters

During the first iteration a segment of data (frame) is presented to the sinusoidal extraction block. The FFT is determined after which the maximum amplitude of the FFT is sought. A rectangular window is used because such a window has the smallest mainlobe width. However the main disadvantage of a rectangular window is the sidelobe attenuation which causes spectral smearing. It is because of this smearing that the extraction process has to be done in an iterative way.

For the maximum found in the FFT a fine search by means of interpolation is made in order to precisely extract the frequency of the underlying sinusoid. When the frequency has been determined the optimal amplitude and phase can be determined by use of linear regression. Finally the sinusoid is generated using the extracted parameters and subtracted from the original segment. This process is repeated so that finally fifty frequencies with accompanying amplitude and phase are extracted per frame. It is noted that this method of extraction doesn't consider whether a spectral peak is actually the result of a sinusoid.

The use of only a short segment length of 720 samples implies that the lower frequencies can't be estimated with high precision. Therefore a multi-scale sinusoidal extraction mechanism has been developed (see Fig. 7). The basic principle of this mechanism is as follows.

First the PCM input signal is fed to an anti-aliasing filter (AAF) after which the signal is downsampled by a factor three (DS3). Now the structure of Figure 5 is applied to 720 samples in the downsampled domain (SE). These samples correspond to three times 720 samples in the original domain. An appropriate segment of downsampled PCM samples that corresponds to the samples in the original domain will therefore have to be selected. This is done in the sinusoidal segmentation unit (SU). This segmentation is shown graphically for all three scales in Figure 6. Note that every segment on the third scale corresponds to multiple segments on

the second scale. Likewise every segment on the second scale corresponds to multiple segments on the first scale. Also note that the segments on the second and third scale are always placed within two transient positions just like the segments (frames) on the first scale.

On the third scale three sinusoids are extracted, on the second scale seven sinusoids and on the first scale forty. When less than three sinusoids are found in the third scale, the second scale may extract seven sinusoids plus what has been left by the third scale. The same applies to the second scale and the first scale.

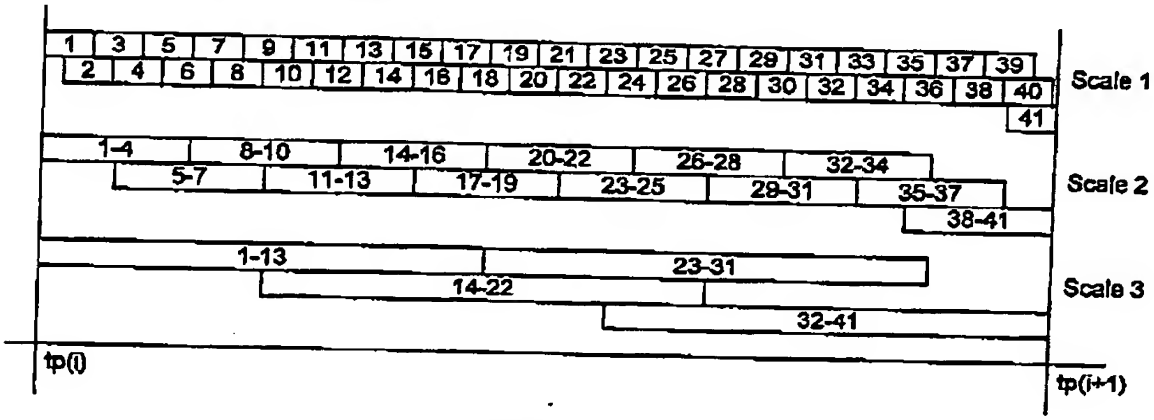

Figure 6: Multi-scale segmentation

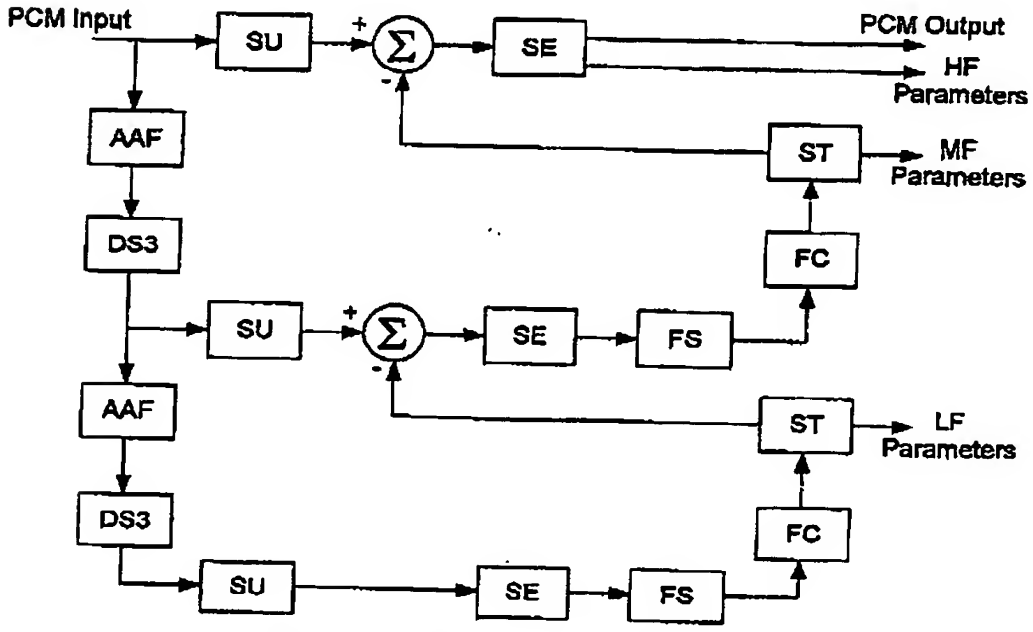

Figure 7: Multi-scale sinusoidal analysis

Because of complexity reasons in both encoder and decoder a description on a single scale is preferable. To come to a single-scale description the estimated parameters must be converted back to the original non-downsampled domain. This is done in a few steps. First of all every scale is band-limited and thus only a limited amount of frequencies may be included. This selection is done in the frequency selection module (FS). For sinusoids that are kept, compensation in amplitude and phase is made for the gain and delay caused by the anti-aliasing filter (FC). Finally the sinusoidal components are converted to the non-downsampled domain by mapping the parameters to the segments (frames) of the previous scale (ST).

Psycho-acoustic model

The key element in removing irrelevancy in audio coding is the psycho-acoustic model. This model tries to describe, given a certain input-signal what parts of that signal can and cannot be perceived by the human auditory system. Psycho-acoustic models can be very comprehensive. They can describe time masking, the masking of parts of the signal over time, frequency masking, the masking of frequency components over each other, stereo masking/unmasking, the masking or unmasking caused by the use of stereo, etc.

The psycho-acoustic model currently used in the SSC encoder only includes a frequency masking model. Every sinusoid can be seen as a frequency component with its own masking ability determined by its power and frequency. All these components together form a so-called masking curve. This is a curve in the frequency domain that describes the total masking ability of all components present. As an example the masking curve of three sinusoids with frequencies of 1000, 2000 and 4000Hz has been calculated (see Fig. 9). Note that the individual masking curves get broader as the frequency gets higher. This effect, as many other laws in psycho-acoustics, are related to the critical band concept. This concept basically states that the human auditory system can be seen as a bank of bandpass filters with different bandwidths. This concept is described by the Equivalent Rectangular Bandwidth scale (ERB) defined as Equation 16:

$$e_f = 21.4\log_{10}\left(\frac{4.37f}{1000}+1\right), \qquad 16$$

where the frequency f is in Hertz and e_f the frequency in erb. The ERB scale is depicted in Figure 8.

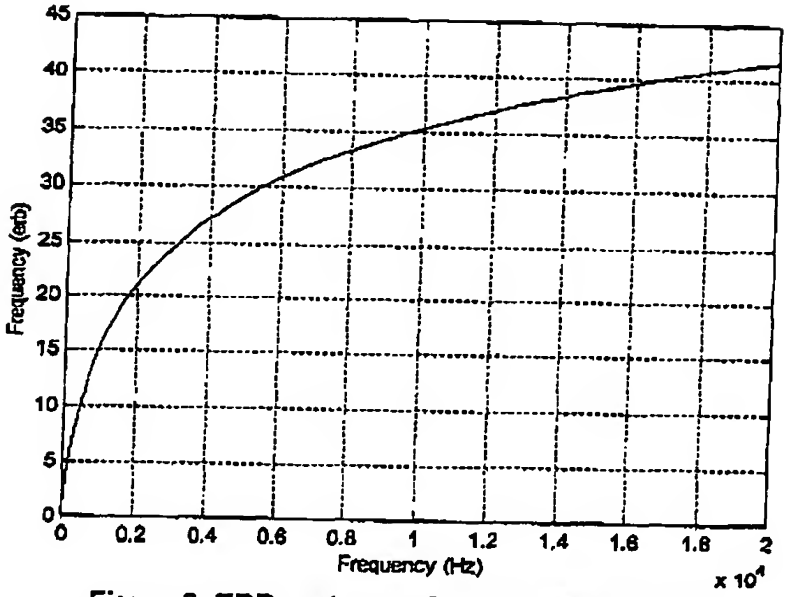

Figure 8: ERB scale as a function of frequency

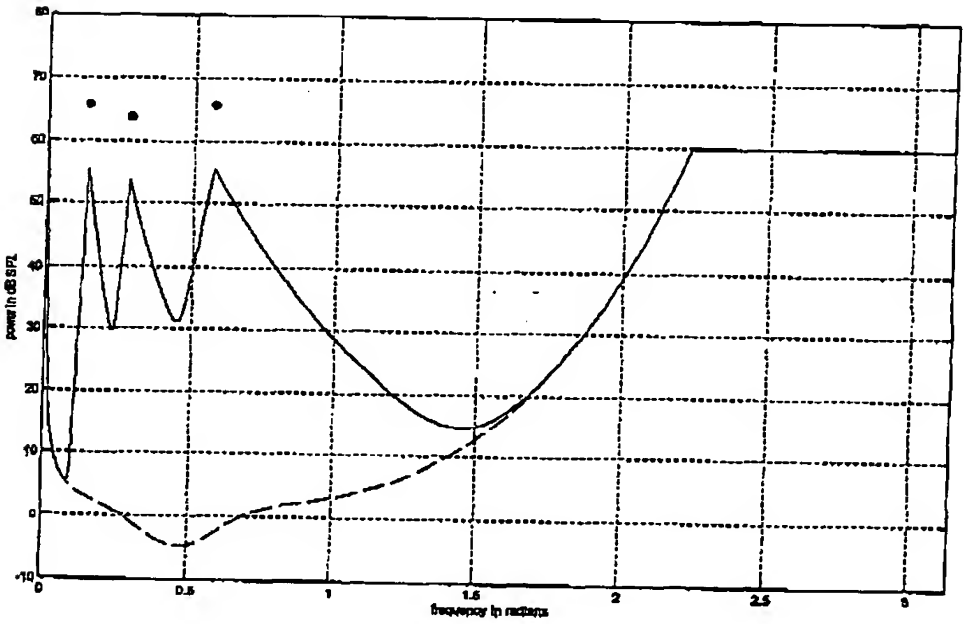

Figure 9: Example of frequency masking of sinusoids

Apart from the individual masking curves the total masking curve is also partially determined by the hearing threshold in quiet. This is a threshold describing how much power a single component should contain in order to be perceived in an absolute sense. The hearing threshold in quiet is shown as an interrupted line, the total masking curve is shown as a solid line in Figure 9.

Tracking

The main function of the tracking algorithm is to improve coding efficiency. The tracking algorithm consists of three steps:

1. Apply tracking: linking of sinusoidal components in time.
2. Phaseless reconstruction: for tracks that have been found only the initial phase has to be transmitted.
3. Removal of non-tracks: tracks that are very short can't be perceived as being a sinusoidal component, they are therefore removed.

The first step of the algorithm tries to link the sinusoidal components that have been found on a frame-to-frame basis. This process is shown graphically in Figure 10. A sinusoidal component will become one of the next four possibilities:

1. Birth: part of a track with only a successor.
2. Continuation: part of a track with predecessor and successor.
3. Death: part of a track with only a predecessor.
4. Non-track: a 'track' consisting of a single frame.

Now the problem arises how to link the sinusoidal components that have been extracted. It is assumed that a track is a slowly varying function of both amplitude and frequency (see Equation 14). Therefore separate cost-functions for both amplitude and frequency have been developed. For the frequency the cost-function is based on the ERB-scale. The reason to do so is that for complex stimuli the relevant events are more or less separated according to the ERB-scale. The cost-function for frequency then becomes (Equation 17):

$$Q^f_{p,q} = \begin{cases} 0 & for \quad \left|e(f_{p,k}) - e(f_{q,k-1})\right| \geq e_{max} \\ 1 - \dfrac{\left|e(f_{p,k}) - e(f_{q,k-1})\right|}{e_{max}} & for \quad \left|e(f_{p,k}) - e(f_{q,k-1})\right| < e_{max}, \end{cases} \qquad 17$$

where $e(f_{p,k})$ denotes the frequency in erb of the p^{th} component in the k^{th} frame and e_{max} the maximally allowed deviation expressed in erb.

For the amplitudes a similar cost-function is used (Equation 18):

$$Q^a_{p,q} = \begin{cases} 0 & for \quad \left|A_{p,k} - A_{q,k-1}\right| \geq A_{max} \\ 1 - \dfrac{\left|A_{p,k} - A_{q,k-1}\right|}{A_{max}} & for \quad \left|A_{p,k} - A_{q,k-1}\right| < A_{max}, \end{cases} \qquad 18$$

where $A_{p,k}$ denotes the amplitude expressed in decibels of the p^{th} sinusoidal component in the k^{th} frame and A_{max} the maximally allowed deviation. The total cost-function now becomes (Equation 19):

$$Q_{p,q} = Q^f_{p,q} Q^a_{p,q}. \qquad 19$$

When for a certain sinusoid p there exists no $Q_{p,q}$ greater than zero it is marked as being the end of a track. If for a certain sinusoid p there exist more than one $Q_{p,q}$ greater than zero sinusoid q is chosen with the largest value of $Q_{p,q}$.

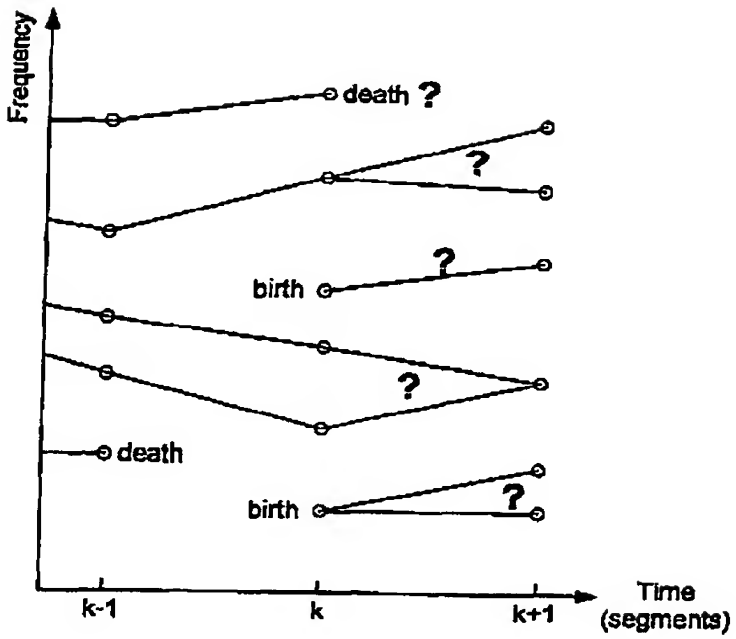

Figure 10: Tracking of sinusoidal components

The second step of the algorithm consists of phaseless reconstruction. Phaseless reconstruction is based on the assumption that the instantaneous phase of a track is a smooth function of time. For two consecutive frames of a track the instantaneous phase is equated as shown below. Assume that the sinusoid in frame *k-1* and frame *k* are described as (Equation 20 and 21):

$$s_{p,k-1}[n] = A_{p,k-1}\cos\left(\omega_{p,k-1}n + \varphi_{p,k-1}\right) \qquad 20$$

$$s_{q,k}[n] = A_{q,k}\cos\left(\omega_{q,k}n + \varphi_{q,k}\right) \qquad 21$$

The equation of the instantaneous phase can be best performed at the middle of the overlapping segments. This is because the synthesis windows are (symmetric) Hanning windows. The overlap is shown graphically in Figure 11 for a segment length of *N=40*. Note that n is defined symmetrically around zero for both Equation 20 as well as 21; shifted time-axes are thus used.

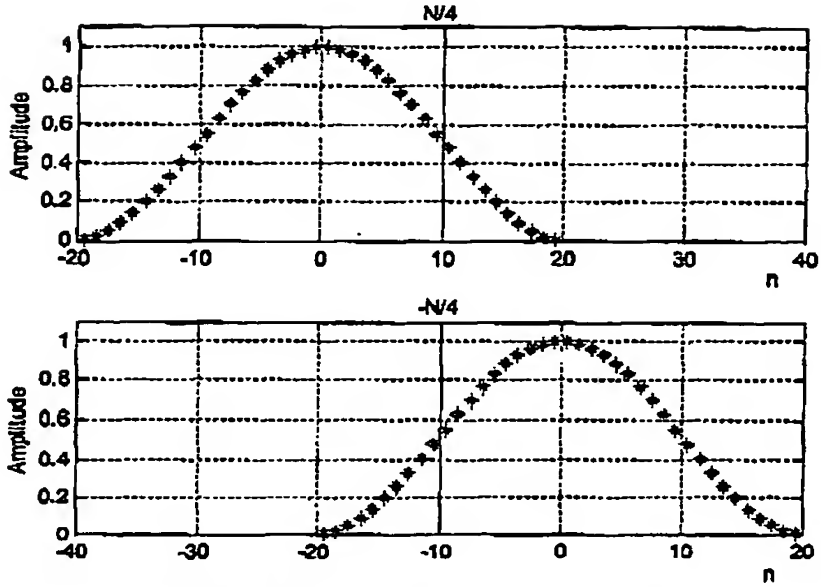

Figure 11: Equation of instantaneous phase at overlapping frames

Equating the instantaneous phase of Equation 20 and 21 at the middle of the overlap of a segment with length *N* then gives (Equation 22):

$$\omega_{p,k-1}\frac{N}{4}+\varphi_{p,k-1}=\omega_{q,k}\frac{-N}{4}+\varphi_{q,k}, \qquad 22$$

which results in (Equation 2.12):

$$\varphi_{q,k}=(\omega_{p,k-1}+\omega_{q,k})\frac{N}{4}+\varphi_{p,k-1}. \qquad 23$$

Equation 23 already indicates that the first step of the tracking algorithm, the linking procedure, has great influence on quality. Erroneous linking of tracks can seriously distort phase relations between tracks.

The final step of the tracking algorithm consists of the removal of short tracks. Tracks that are shorter than five sinusoidal periods cannot be perceived by the human auditory system as being a tonal component. Such tracks are therefore deleted.

Coding of sinusoidal components

The coding of the sinusoidal components consists of:

1. Quantisation of parameters.
2. Sort data in births and continuations.
3. Sort births in frequency.
4. Sort continuations in frequency (where deaths are also seen as continuations).
5. Apply absolute and differential coding.

The quantisation of the parameters is done according to the same rules as the sinusoids in the transient code.

In order to efficiently code the parameters and the tracking information the matrices containing the sinusoidal components' frequency, amplitude and phase are sorted. This is shown graphically in Figure 12. The left matrix shows how the information is stored before

sorting; the right matrix shows how the information is stored after sorting. Note that for both matrices the only information needed to pick the right index of the next set of parameters belonging to a track is whether or not a sinusoidal component is continued.

For the first birth in a frame the amplitude and frequency are coded absolutely. The amplitude and frequency for all other births within a frame are coded differentially to their predecessor in the frequency domain (vertically in Fig. 12). For continuations the amplitude and frequency is coded differentially in the time domain (horizontally in Fig. 12).

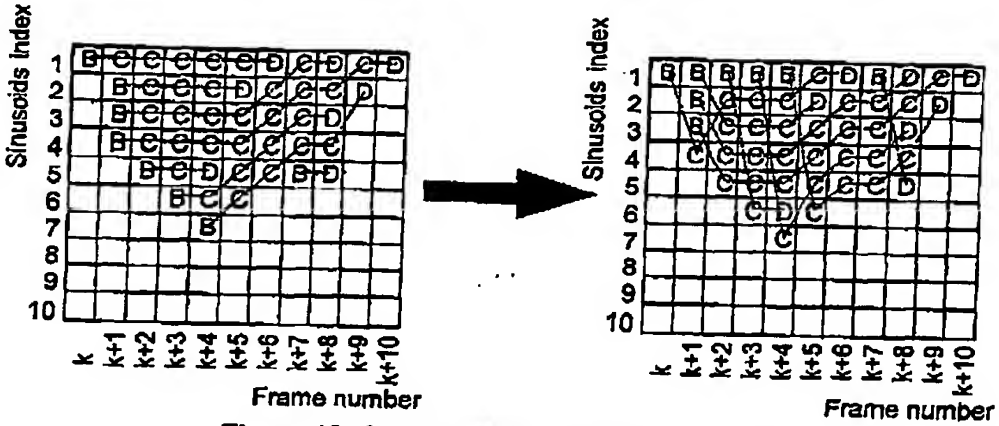

Figure 12: Sorting of births and continuations

Sinusoidal Synthesis

The synthesis of the sinusoidal components differs from the analysis only in the windowing that is used. In the analysis section overlapping frames were analysed as (Equation 24):

$$s_k[n] = \sum_p A_{p,k} \cos\left(\omega_{p,k} n + \varphi_{p,k}\right), \qquad 24$$

where p denotes sinusoid index and k the frame index.
In the synthesis section the overlapping frames are synthesised as (Equation 25):

$$s_k[n] = h[n] * \sum_p A_{p,k} \cos\left(\omega_{p,k} n + \varphi_{p,k}\right), \qquad 25$$

where $h[n]$ denotes the window function and * denotes convolution. The windows that are used are amplitude complementary. Three types of windows are used during synthesis:

1. Normal window, defined as a Hanning window (Equation 26):

$$h[n] = \frac{1}{2} + \frac{1}{2}\cos\left(\frac{2\pi}{N}\left(n - \frac{N-1}{2}\right)\right) \quad for \quad 0 \le n \le N-1, \qquad 26$$

where N denotes the frame length.

2. Start window, defined as half rectangular and half a Hanning window (Eq. 27):

$$h[n]=\begin{cases}1 & for \quad 0\le n\le\dfrac{N-1}{2}\\ \dfrac{1}{2}+\dfrac{1}{2}\cos\left(\dfrac{2\pi}{N}\left(n-\dfrac{N-1}{2}\right)\right) & for \quad \dfrac{N-1}{2}<n\le N-1.\end{cases} \qquad 27$$

3. Stop window, defined as half a Hanning window and for the rest a rectangular window (Equation 28):

$$h[n]=\begin{cases}\dfrac{1}{2}+\dfrac{1}{2}\cos\left(\dfrac{2\pi}{N}\left(n-\dfrac{N-1}{2}\right)\right) & for \quad 0\le n\le\dfrac{N-1}{2}\\ 1 & for \quad \dfrac{N-1}{2}<n\le T-N-1,\end{cases} \qquad 28$$

where T denotes the length of the frame needed.

Only the frames at the edges of transients use start- and stop-windows (frames 1 and 41 in Figure 6), all others use normal windows (frames 2 until 40 in Figure 6). A typical windowing sequence is depicted in Figure 13.

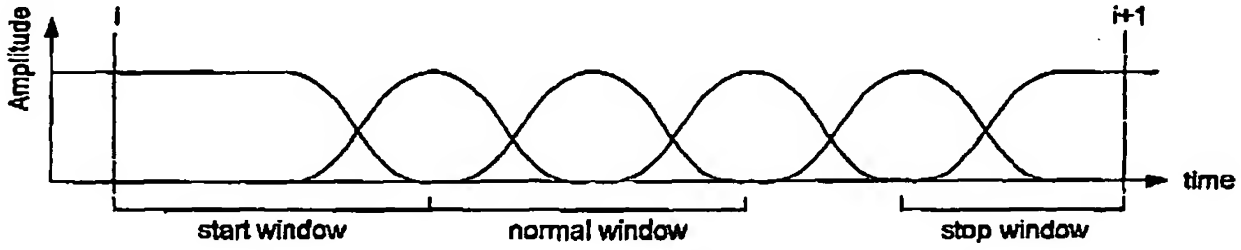

Figure 13: Typical synthesis windowing sequence

QUANTISATION AND CODING

Until now the focus was mainly on the extraction and processing of the parameters towards an efficient representation. It was also shown how irrelevancy is exploited within the SSC coder. To further improve coding efficiency not only irrelevancy should be exploited but also the redundancy.

To efficiently encode all the extracted and processed parameters into a bit-stream three steps are required:

1. Pre-processing (sorting),
2. Quantisation,
3. Entropy coding.

First of all the parameters need to be pre-processed. This is mainly a sorting process in which the different parameters are grouped together to form sets that can each be represented efficiently. For example the sorting of births and continuations as described above belongs to the pre-processing stage, but also the sorting of sinusoidal components by frequency in the transient module.

In the second stage, quantisation is applied to the sorted parameters. The parameters must be quantised in such a manner that after decoding *just* no differences can be perceived before and after quantisation. For e.g. the amplitudes of the sinusoids logarithmic quantisation can be applied.

Finally entropy coding is applied to the representation levels, which represent the quantised levels. The most common form of entropy coding is Huffman coding. This is a constant to variable wordlength entropy coding technique. The main advantage of Huffman coding over other entropy coders is the low complexity: both encoding and decoding can be performed by table look-up. Furthermore these tables are easily constructed.

A general diagram for coding audio parameters using Huffman coding is depicted in Figure 14.

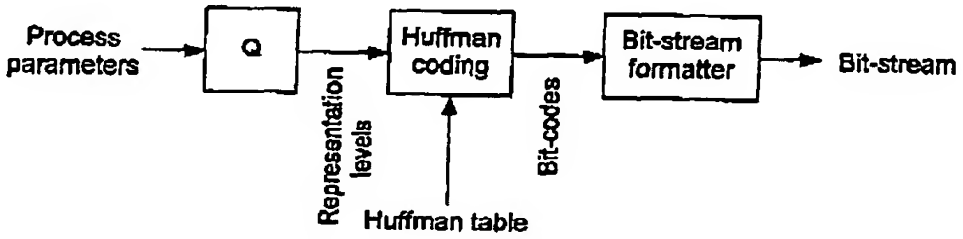

Figure 14: General block-diagram for efficiently encoding of parameters

One of the main problems with SSC proved to be the quality scalability. From a certain bit-rate on the quality doesn't increase anymore. This is shown graphically in Figure 15.

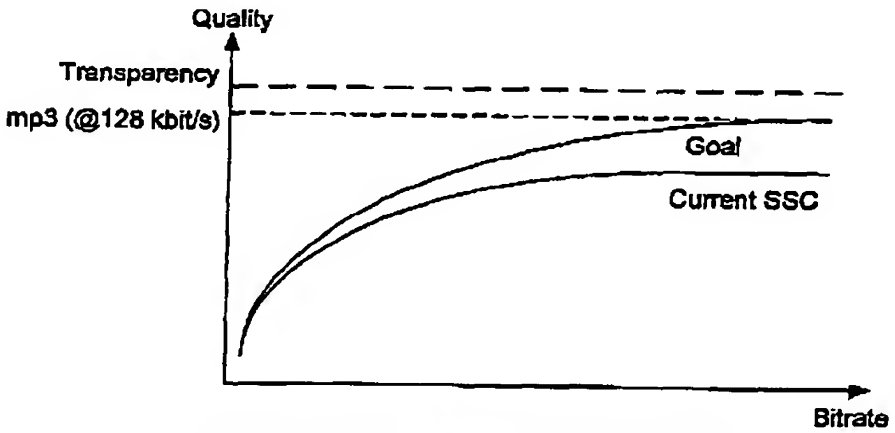

Figure 15: Quality scalability of the SSC coder

A goal of the invention is to address this problem.

TRACKING ALGORITHM

The tracking algorithm is both essential in terms of bit-rate reduction and perceived audio quality. In principle the linking of individual sinusoids as it is, only has influence on the bit-rate. However, since inextricably bound up with the phase-less reconstruction within a track, the quality is also affected. A phase-less reconstruction and the removal of non-tracks cause a significant perceptual loss of quality.

In this chapter a new approach to the tracking algorithm is presented, based on the information obtained from the second order polynomial description of sinusoids. This results in a decreased loss of quality. Also adjustments to the algorithm that removes non-tracks are described.

LINKING OF SINUSOIDAL COMPONENTS

A goal of the tracking algorithm is to link individual sinusoids between consecutive frames. This has two purposes. First of all, a track is an efficient representation of a sinusoid over time. The frequency and amplitude can e.g. be coded differentially. Secondly, if tracks represent the true course of the sinusoids over time, the phase information becomes redundant and can therefore be discarded. If however incorrect links have been made, this will cause audible discontinuities.

When observing spectrograms, the process of linking seems pretty straightforward. In practice however some problems occur. To illustrate what problems can occur a synthetic signal is generated (Equation 29):

$$s[n] = \sum_{p=1}^{P} \frac{A}{2p-1} \sin\left((2p-1)(\omega + \omega_d \sin(n\omega_m))n\right), \qquad 29$$

where p indexes the p^{th} sinusoid, P describes the total number of sinusoids, A is the amplitude of the fundamental harmonic, ω the frequency of the fundamental harmonic, ω_d the modulation depth and ω_m the modulation frequency. Figure 16 shows a part of the spectrogram of this signal for $A = 1000$, $P = 30$, $\omega = 100F_s/2\pi$, $\omega_d = 1.5F_s/2\pi$ and $\omega_m = F_s/2\pi$.

The signal as described in Equation 29 basically describes an approximation of a frequency modulated square-wave by thirty sinusoids. As can be seen in Figure 29, the amount of frequency modulation increases over time. Visually the thirty spectral lines, each representing a single sinusoid, are easily followed. The tracking of this signal thus appears to be straightforward. Two types of problems however occur:

1. Wrong connections: two sinusoids of consecutive frames get connected while they should not have been.
2. Non-connection: two sinusoids of consecutive frames don't get connected while they should have been.

Figure 30 shows how the current tracking mechanism, using Equations 17, 18 and 19, connects the extracted sinusoids for a small part of the signal in Figure 16. The solid lines describe tracks, where the crosses denote births (beginnings) and deaths (endings) of tracks.

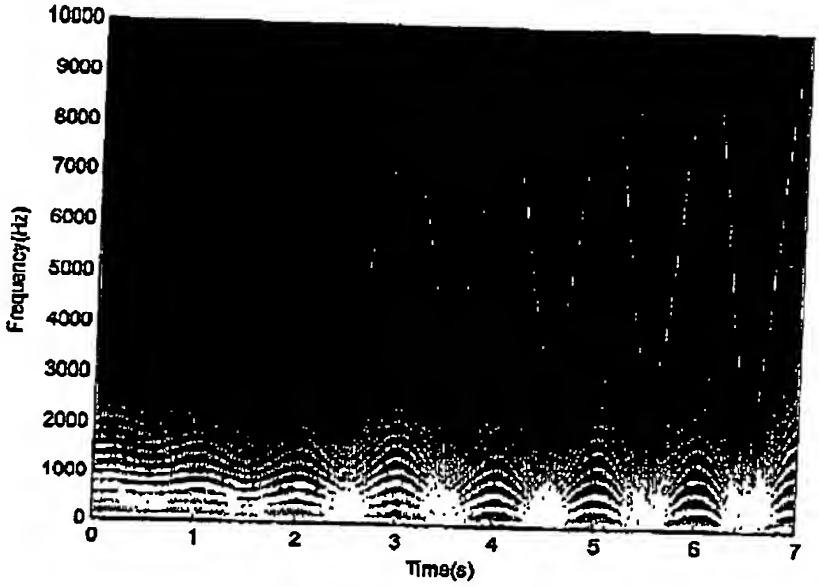

Figure 16: Spectrogram of synthetic signal

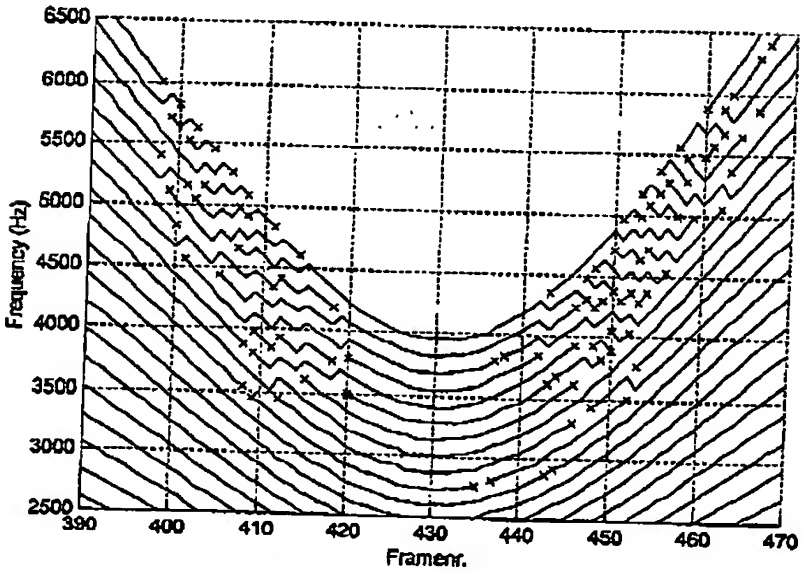

Figure 17: Tracking of sinusoidal components for synthetic signal

In Figure 17 the problems described above are clearly visible. Especially high frequency variations cause erroneous linking of sinusoids. This can be explained by the cost-function (Equation 19). One of the partial cost-functions consists of a criterion, based on the difference in frequency of sinusoids in subsequent frames. This causes the linking mechanism to have a tendency to go straight on. This is depicted in Figure 18: two subsequent frames are shown as planes in a three-dimensional environment. The *x*-axis describes time, the *y*-axis the amplitude and the *z*-axis (going into the paper) describes frequency. The dots on these planes denote sinusoidal components. For the p^{th} sinusoid in frame k the cost-function is visualised as a circle. Sinusoids that fall outside of the circle will not be linked to the sinusoid. For all sinusoids

that fall within the circle the sinusoid that is closest to the centre of the circle will be chosen as link. In Figure 18 this is depicted using grey values, where a darker grey gives preference over a lighter grey.

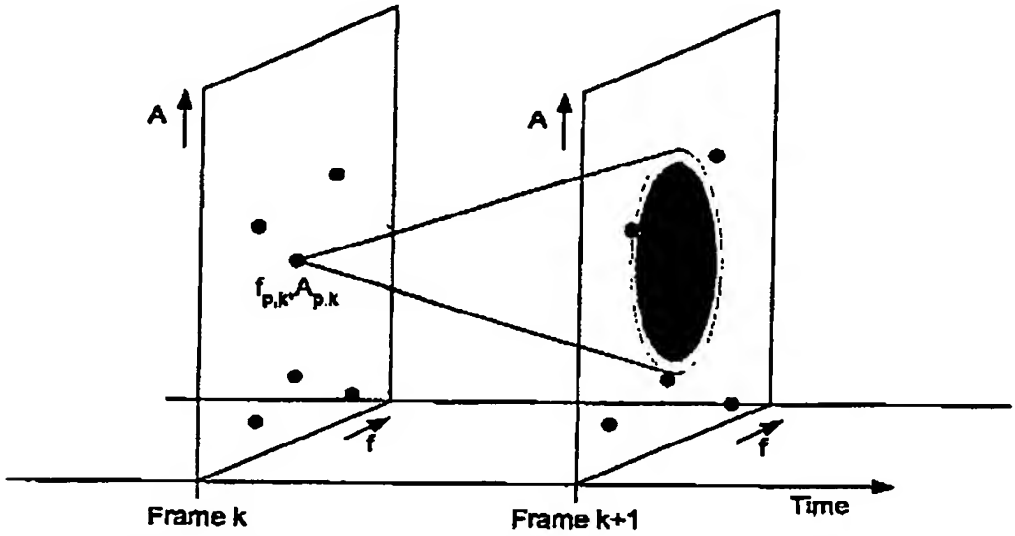

Figure 18: 'Straight ahead'-property of current linking mechanism

Besides the 'straight ahead'-property, Figure 18 also indicates another weakness of the current linking mechanism: time information isn't taken into account during the linking process. The analysis of sinusoidal components took place using 50% overlap. This means that on an average basis the matching of sinusoids is at best at half the overlap. Ideally the linking should thus be performed there. The only time-information that is available, the phases of the sinusoids, is currently not exploited.

The linking mechanism can be drastically improved using the information obtained from the second order polynomial sinusoid description. For this purpose a sinusoid is written once again as the real part of three complex vectors (Equation 30):

$$s_p[n] = \Re\left\{\left(a + bn + cn^2\right)e^{j\omega n}\right\} \qquad 30$$

This description of a sinusoid is depicted in Figure 19. In this figure the three complex vectors $ae^{j\omega n}$, $bne^{j\omega n}$ and $cn^2e^{j\omega n}$ of Equation 30 are shown on the complex plane for two consecutive time instances. The mapping on the real axis of the total vector, given as the sum of the three complex vectors, describes $s_p[n]$. Note that even though each complex vector rotates with frequency ω, the total vector, given as the sum of the three complex vectors, does not necessarily rotate with this same frequency ω. This figure thus also proves that frequency variation can be handled using the three complex vectors.

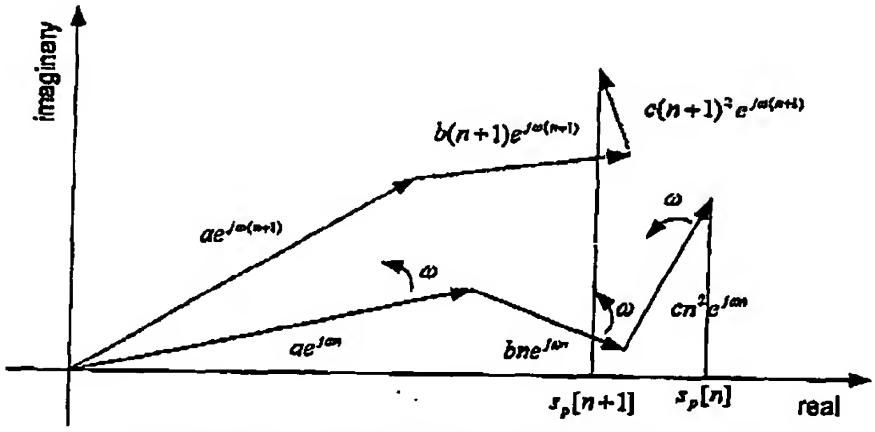

Figure 19: Sinusoid as a sum of complex vectors

The three complex vectors of Equation 30 can also be seen as a single complex vector consisting of the sum of these three vectors.

$$s_p[n] = \Re\{C(n)\},$$
$$= \Re\{R(n)e^{jp(n)}\}, \qquad 31$$

where $C(n)$ denotes the complex vector representing the whole signal, $R(n)$ the envelope function and $p(n)$ the instantaneous phase function. As an example a sinusoid is shown graphically as a single complex vector in Figure 20.

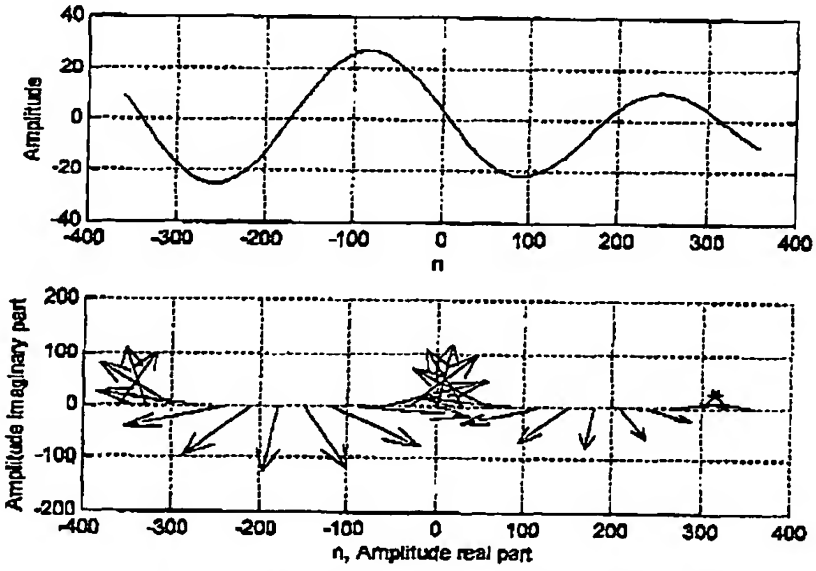

Figure 20: Sinusoid as a single complex vector

The upper figure of Figure 20 shows the waveform as a function of time. The lower figure shows complex samples of this same waveform denoted using arrows. Each arrow describes a sample of the waveform in its complex form (see Equation 31). Figure 20 tries to indicate that the complex description using the parameters: instantaneous phase,

instantaneous frequency and instantaneous amplitude (envelope) is no different from the waveform description. The shape of the waveform of a single sinusoid is thus fully described by Equation 31. The instantaneous amplitude $R(n)$ is the time-dependent equivalent of the amplitude A, the instantaneous phase $p(n)$ is the time-dependent equivalent of the phase φ and the instantaneous frequency $dp(n)/dn$ is the time-dependent equivalent of the frequency ω. The second-order polynomial thus gives extra information on how the amplitude, phase and frequency vary over time, within a segment. This extra information can be used to improve the linking mechanism.

In the case a sinusoid is part of a track, there should be a close match of waveforms of sinusoids in consecutive frames. Figure 21 shows two waveforms of overlapping frames with such a close match at the overlap region. A possible linking criterion account for the amount of matching found between sinusoids at the overlap region. Equation 31 is ideal for such a purpose because it describes the three independent complex vectors of a single sinusoid as one. The question now remains how to define a matching criterion on $C(n)$.

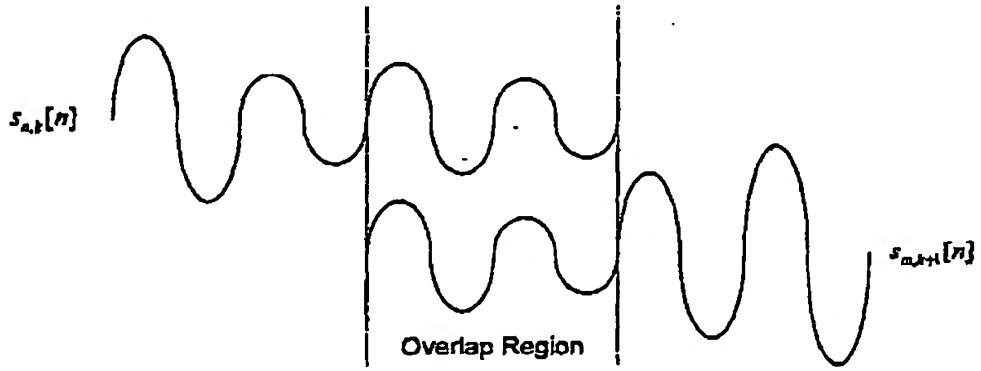

Figure 21: Waveform matching in overlapping frames

Now, fully analogous to the current linking mechanism, the following cost-function has been defined (Equation 32):

$$Q_{p,q} = Q^{R}_{p,q} Q^{F}_{p,q} Q^{\Omega}_{p,q}. \qquad 32$$

where the subscript p denotes the p^{th} sinusoid of frame k and subscript q denotes the q^{th} sinusoid of frame $k-1$. When normal windows are applied, the cost-function is applied at the middle of two overlapping segments. When start- or stop-windows are used, the cost-function is applied at the edges. This is shown graphically in Figure 22. So, instead of the current cost-function that makes use of global values, now instantaneous values are used.

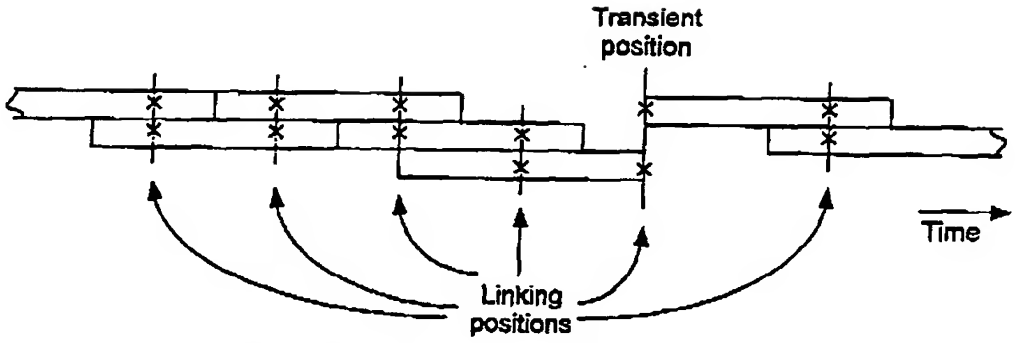

Figure 22: Linking positions according to segmentation

The cost-function for the *instantaneous* amplitude has been defined as (Equation 33):

$$Q_{p,q}^{R} = \begin{cases} 0 & for \quad \left|R_{p,k} - R_{q,k-1}\right| \geq R_{\max} \\ 1 - \dfrac{\left|R_{p,k} - R_{q,k-1}\right|}{R_{\max}} & for \quad \left|R_{p,k} - R_{q,k-1}\right| < R_{\max}, \end{cases} \qquad 33$$

where R denotes the instantaneous amplitude expressed in decibels and $R_{\max}$ denotes the maximally allowed deviation in decibels. Both are expressed in decibels to match the human auditory system.

The cost-function for the instantaneous frequency is defined as (Equation 34):

$$Q_{p,q}^{\Omega} = \begin{cases} 0 & for \quad \left|e(\Omega_{p,k}) - e(\Omega_{q,k-1})\right| \geq e(\Omega_{\max}) \\ 1 - \dfrac{\left|e(\Omega_{p,k}) - e(\Omega_{q,k-1})\right|}{e(\Omega_{\max})} & for \quad \left|e(\Omega_{p,k}) - e(\Omega_{q,k-1})\right| < e(\Omega_{\max}), \end{cases} \qquad 34$$

where $e(.)$ denotes the frequency expressed in ERB. The instantaneous frequency is approximated using a first order difference (Equation 35):

$$\begin{aligned} \Omega_{p,k} &= p_{p,k}\left(n_{overlap,k} + 1\right) - p_{p,k}\left(n_{overlap,k}\right), \\ \Omega_{q,k-1} &= p_{q,k-1}\left(n_{overlap,k-1} + 1\right) - p_{q,k-1}\left(n_{overlap,k-1}\right), \end{aligned} \qquad 35$$

where $n_{overlap,k}$ denotes the overlap position of a frame k at the left-hand side, $n_{overlap,k-1}$ denotes the overlap position of a frame $k-1$ at the right-hand side as shown in Figure 22, and $p_{p,k}(n)$ the instantaneous phase of the p^{th} sinusoid in frame k at position n.

The cost-function for the instantaneous phase is defined as (Equation 36):

$$Q_{p,q}^{p} = \begin{cases} 0 & for \quad \left|p_{p,k} - p_{q,k-1}\right| \geq p_{\max} \\ 1 - \dfrac{\left|p_{p,k} - p_{q,k-1}\right|}{p_{\max}} & for \quad \left|p_{p,k} - p_{q,k-1}\right| < p_{\max}, \end{cases} \qquad 36$$

where $p_{p,k}$ and $p_{q,k-1}$ are defined as (Equation 37):

$$p_{p,k} = p_{p,k}\left(n_{overlap,k}\right)$$
$$p_{q,k-1} = p_{q,k-1}\left(n_{overlap,k-1}\right) \qquad 37$$

where $n_{overlap,k}$ and $n_{overlap,k-1}$ are defined as described above.

Experiments using the partial cost-functions as described above showed that the partial cost-function for the instantaneous frequency (Equation 34) sometimes behaved unpredictably. Some further research showed that this happened especially at areas where the envelope-function is close to zero. Figure 23 shows an example of such behaviour. The upper figure shows an amplitude-modulated sinusoid as a function of time. The lower figure shows the instantaneous frequency that belongs to the waveform of the upper figure. As can be seen quite clearly the instantaneous frequency around position zero becomes negative. In itself this isn't a problem; both sinusoids that are to be linked will show such behaviour. Some experiments however showed that the matching of waveforms at such critical areas in general isn't very good. This especially affects the instantaneous frequency. Therefore the cost-function of the instantaneous frequency was replaced with the cost-function for the (modulation) frequency ω.

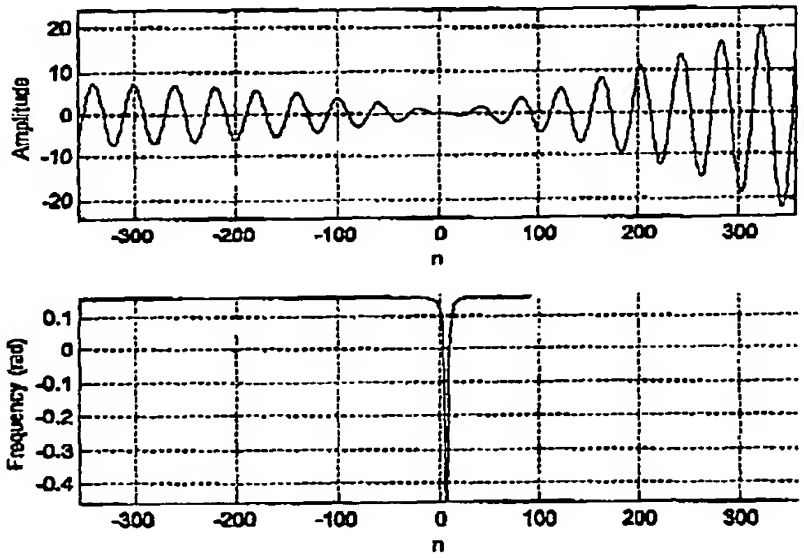

Figure 23: Amplitude modulated sinusoids and its instantaneous frequency

The cost-function for the frequency then becomes identical to the current cost-function for the frequency (Equation 38):

$$Q^{\omega}_{p,q} = \begin{cases} 0 & for \quad \left|e(\omega_{p,k}) - e(\omega_{q,k-1})\right| \geq e(\omega_{max}) \\ 1 - \dfrac{\left|e(\omega_{p,k}) - e(\omega_{q,k-1})\right|}{e(\omega_{max})} & for \quad \left|e(\omega_{p,k}) - e(\omega_{q,k-1})\right| < e(\omega_{max}). \end{cases} \qquad 38$$

The partial cost-functions of Equation 33, 36 and 38 contain respectively the thresholds R_{max}, p_{max} and $e(\omega_{max})$, which are still to be determined. At first this has been done in such a way, that for almost every track that is clearly visible in a spectrogram, links are indeed made.

Secondly an optimisation towards quality has been made. For this purpose the values of R_{max}, p_{max} and $e(\omega_{max})$ have been slightly adjusted for an optimum between quality and amount of links. A smaller amount of links means that more births will occur. Every birth is coded including its original phase. This in comparison to a continuation costs more bits. However, the quality in general will increase, but so will the bit-rate. A balance has been sought between the amount of links and the perceived audio quality. This led to the following values: $e(\omega_{max}) = 0.5$ erb, $p_{max} = 1/3\pi$ radians and $R_{max} = 12$ dB.

As an example, Figure 17 is depicted once again, but now with improved linking as described above (see Figure 24). The solid lines once again denote tracks that have been formed; the crosses denote births and deaths of tracks. When compared to Figure 17 the improvement is quite clear. However still some erroneous links are made and sometimes no connection is made at all.

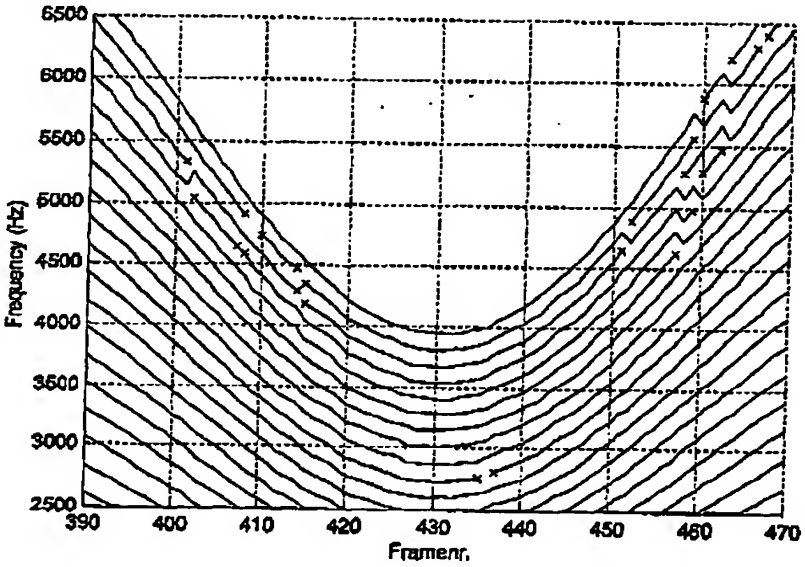

Figure 24: Linking on synthetic signal of Equation 19

CONTINUATION OF PHASE

Above following equation was derived for the continuation of the phase (Equation 39):

$$\varphi_{q,k} = (\omega_{p,k-1} + \omega_{q,k})\frac{N}{4} + \varphi_{p,k-1}. \qquad 39$$

This equation was derived from the concept of constant sinusoids. As seen in Chapter 4 to a certain extent frequency variation can also be modelled. As an approximation the p^{th} sinusoid in frame k then is described as (Equation 40):

$$s_{p,k}(n) = A_{p,k}\cos\left(\left(\omega_{p,k} + \Delta\omega_{p,k}n\right)n + \varphi_{p,k}\right) \qquad 40$$

where $\Delta\omega_{p,k}$ describes linear frequency variation. The q^{th} sinusoid in frame $k-1$ can then be described on the same time-axis as Equation 40 as:

$$s_{q,k-1}(n) = A_{q,k-1}\cos\left(\left(\omega_{q,k-1} + \Delta\omega_{q,k-1}\left(n+\frac{N}{2}\right)\right)\left(n+\frac{N}{2}\right) + \varphi_{q,k-1}\right) \quad 41$$

Suppose that both sinusoids of Equation 40 and Equation 41 have been linked together. To obtain a smooth transition between the two sinusoids, the instantaneous phases must be equated (Equation 42):

$$\left(\omega_{p,k} + \Delta\omega_{p,k}n\right)n + \varphi_{p,k} = \left(\omega_{q,k-1} + \Delta\omega_{q,k-1}\left(n+\frac{N}{2}\right)\right)\left(n+\frac{N}{2}\right) + \varphi_{q,k-1}. \quad 42$$

Now suppose the two segments will be linked in the middle of the overlap ($n=-N/4$), as is the general case (Equation 43):

$$-\left(\omega_{p,k} - \Delta\omega_{p,k}\frac{N}{4}\right)\frac{N}{4} + \varphi_{p,k} = \left(\omega_{q,k-1} + \Delta\omega_{q,k-1}\left(-\frac{N}{4}+\frac{N}{2}\right)\right)\left(-\frac{N}{4}+\frac{N}{2}\right) + \varphi_{q,k-1}. \quad 43$$

Using Equation 43 the following can be derived (Equation 44):

$$\varphi_{p,k} = \frac{N}{4}\left(\omega_{p,k} + \omega_{q,k-1}\right) + \frac{N^2}{16}\left(\Delta\omega_{q,k-1} - \Delta\omega_{p,k}\right) + \varphi_{q,k-1}. \quad 44$$

For the two terms $\Delta\omega_{q,k-1}$ and $\Delta\omega_{p,k}$ different approximations can be used:

- Forward and backward differences (Equation 45):

$$\Delta\omega^{F}_{p,k} = 2\frac{\omega_{r,k+1} - \omega_{p,k}}{N}, \qquad \Delta\omega^{B}_{p,k} = 2\frac{\omega_{p,k} - \omega_{q,k-1}}{N},$$

$$\Delta\omega^{F}_{q,k-1} = 2\frac{\omega_{p,k} - \omega_{q,k-1}}{N}, \qquad \Delta\omega^{B}_{q,k-1} = 2\frac{\omega_{q,k-1} - \omega_{s,k-2}}{N}. \quad 45$$

- Approximations based on second order polynomials

In this embodiment, the approximations based on the second order polynomials are of no use for the decoder because the second order polynomials will not be included in the bit-stream.

Different combinations of the forward and backward differences however might give better results. Therefore they have been evaluated using informal listening tests. For $\Delta\omega_{q,k-1}$ and $\Delta\omega_{p,k}$ respectively the following combinations have been evaluated: forward-forward, forward-backward, backward-forward and backward-backward. Note that the combination forward-backward leads to Equation 40. Also note that the forward difference of $\Delta\omega_{p,k}$ needs future information of the track while the backward difference of $\Delta\omega_{q,k-1}$ needs history of the track. Both differences can thus only be applied on a limited area of a track.

In informal listening tests signals where generated using the four combinations as described above. None of the combinations proved superior to another. Therefore equation 40 was preferred over any other combination.

Without further discussion, in this chapter the assumption has been made that the continuation of the phase should always be performed from the birth of a track. There are however reasons to choose another initialisation point within a track:

- Most sinusoids will not exactly start at the segmentation boundaries. For such a sinusoid, the parameters can be estimated incorrectly. The same applies to the end of a sinusoid. This, in contrast to the current continuation of phase, indicates that a distance is to be kept from the beginning and the end of a track.
- The accumulation of errors will be smaller when the continuation is started from the middle.
- The encoder delay becomes larger when the continuation of the phase is performed from a point away from the start of a track. So, a point close to the beginning of a track is preferred. Note that for the decoder no such problems occur. The original phase can be calculated all the way back to the birth of a track. This is shown graphically in Figure 25. Just like Equation 39, the continuation of the phase one frame forward in time, the continuation of the phase one frame backwards in time can be calculated. So when a whole track has been determined starting with the original phase at frame *k*, the continued phase at frame *k-1* can be calculated. From this frame, the continued phase at frame *k-2* can be calculated, etc.

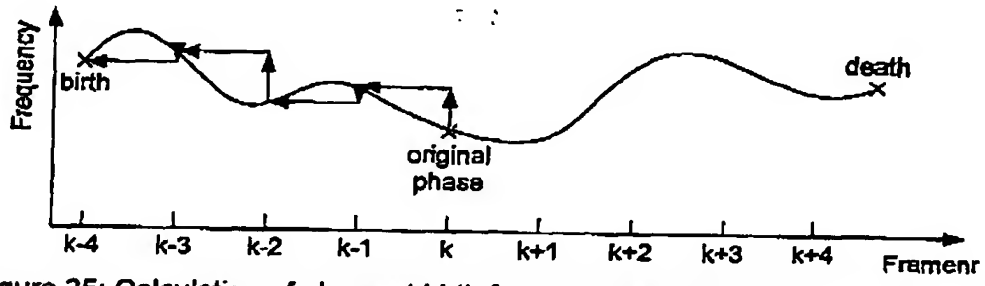

Figure 25: Calculation of phase at birth from original phase at arbitrary position

Another set of informal listening tests has been performed in order to assess the optimal place within a track of length L for the original phase. The following configurations have been tested:

1. Original phase at birth of a track: at $L=1$, just like the current continuation.
2. Original phase placed in the middle of a track: at $\lceil L/2 \rceil$.
3. For tracks with $L>10$: original phase placed at $L=5$, for shorter tracks at $L=1$.
4. For tracks with $L>5$: original phase placed at $L=3$, for shorter tracks at $L=1$.
5. For all tracks with $L>1$: original phase placed at $L=2$, else at $L=1$.

Although the perceptual differences where quite small, configuration 2,3,4 and 5 all outperformed configuration 1, as expected. For the others no preference could be given, therefore configuration 5 is preferred because of its short delay.

EVALUATION OF TRACKING MECHANISM

All in all the quality can be drastically improved using the current tracking mechanism. However (near-)transparency doesn't seem feasible using constraints of the same order as in

the improved linking mechanism described above. Stricter linking-constraints will however increase the bit-rate which is unwanted.

REMOVAL OF NON-TRACKS

Non-tracks are sinusoidal components that have not been linked, i.e. their birth and death takes place in the same frame. Such components are expensive in terms of bit-rate, while perceptually they are often irrelevant. In order to achieve low bit-rates, removal therefore is necessary.

From the field of psycho-acoustics there is not much knowledge on perceptual relevance of non-tracks. It is however known that sinusoids consisting of less than five periods aren't perceived as being sinusoidal. Therefore in the current removal-algorithm all sinusoids with a track-length of one and all tracks consisting of less than five sinusoidal periods are discarded.

Removal of non-tracks caused a perceptual loss of quality, especially around transient positions. This in essence means that too much sinusoidal elements are discarded. Two main reasons can be given:

- Transients typically contain of a lot of short sinusoids. This can be seen in spectrograms where transients typically show up as vertical lines.
- As described above, sinusoids consisting of less than five periods aren't perceived as being sinusoidal. This however doesn't mean they may be left out without consequences.

In order to improve the quality of the removal-algorithm a concession towards the bit-rate will have to be made. This is done by means of the inclusion of a psycho-acoustic model. First of all on a frame-to-frame basis the masking curve is calculated with a low in-band masking. This is shown in Figure 26 and 27.

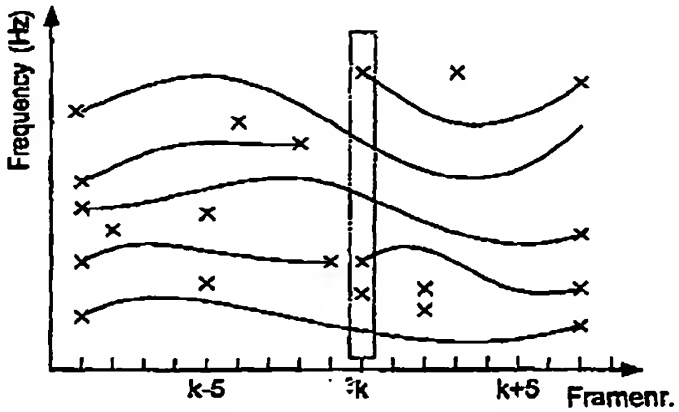

Figure 26: Selection of one frame for psycho-acoustic model

In Figure 26 the solid lines denote tracks; the crosses denote births and deaths of frames. Crosses that aren't at the beginning or end of a line denote the non-tracks. Figure 27 shows the masking curve of the sinusoidal components of the selected frame of Figure 26. Two sinusoidal components fall under the masking curve. However only the sinusoidal component that is a non-track is removed.

Informal listening tests showed that this technique could be applied with an in-band masking of up to 0 dB without any loss of quality. This however means that especially around transients a lot more information (sinusoids) will have to be included which will lead to a higher bit-rate. This is described, among other subjects, in the next chapter.

Furthermore, the tracks that contain less than five periods are still removed. Informal listening tests showed that this did not further degrade the quality.

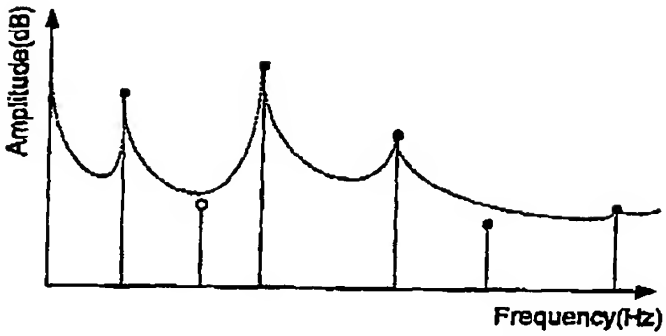

Figure 27: Frequency masking curve of sinusoidal components

All in all the removal of the non-tracks had to be restricted in order to preserve the audio quality. In other words less sinusoids are removed, and thus more sinusoids have to be encoded. This will always lead to a higher bit-rate.

CONCLUSIONS

Visual observation of tracks of synthetic signals, indicates that the operation of the linking algorithm seems to improve drastically when using information obtained from the 2nd order polynomial information. Experiments also showed that the use of continuous phase, with the current constraints in the linking mechanism, still does not deliver near-transparent quality. Furthermore, as shown in the previous chapter, the improvement of the tracking algorithm costs about 2 kbit/s extra bit-rate. Correct tracking thus costs extra bit-rate.

The 2nd order polynomial description in the sinusoidal tracking may advantageously be applied in an SSC encoder.

CLAIMS:

1. A parametric coding method, wherein sinusoidal tracks are formed by estimation of sinusoidal parameters per segment and a linking procedure which establishes a similarity between estimated signal components in two subsequent segments, wherein in the linking procedure a linking criterion is used which is based on an error measure of instantaneous signal parameters at a seam between the two subsequent frames.

2. A parametric encoder comprising means for forming sinusoidal tracks by estimation of sinusoidal parameters per segment and a linking procedure which establishes a similarity between estimated signal components in two subsequent segments, wherein in the linking procedure a linking criterion is used which is based on an error measure of instantaneous signal parameters at a seam between the two subsequent frames.

THIS PAGE BLANK (USPTO)